

# Analysis of Isotopically Depleted Proteins Derived from *Escherichia coli* and *Caenorhabditis elegans* Cell Lines by Liquid Chromatography 21 T Fourier Transform-Ion Cyclotron Resonance Mass Spectrometry

Zeljka Popovic, Lissa C. Anderson, Xuepei Zhang, David S. Butcher, Greg T. Blakney, Roman A. Zubarev, and Alan G. Marshall\*




Cite This: *J. Am. Soc. Mass Spectrom.* 2023, 34, 137–144



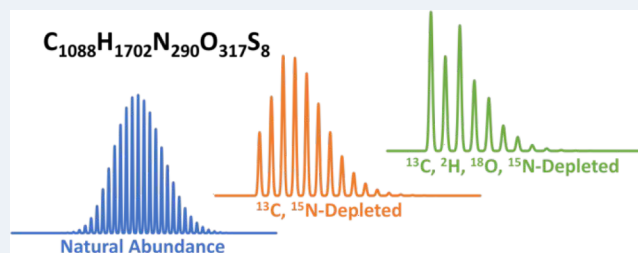
Read Online

ACCESS |

 Metrics & More

 Article Recommendations

**ABSTRACT:** Protein mass measurement by mass spectrometry is complicated by wide isotopic distributions that result from incorporation of heavy isotopes of C, H, N, O, and S, thereby limiting signal-to-noise ratio (SNR) and accurate intact mass determination, particularly for larger proteins [Fenselau et al. *Anal. Chem.* 1983, 55 (2), 353–356]. Observation of the monoisotopic mass-to-charge ratio ( $m/z$ ) is the simplest and most accurate way to determine intact protein mass, but as mass increases, the relative abundance of the monoisotopic peak becomes so low that it is often undetectable. Here, we used an isotopically depleted growth medium to culture bacterial cells (*Escherichia coli*), resulting in isotopically depleted proteins. Isotopically depleted proteins show increased sequence coverage, mass measurement accuracy, and increased S/N of the monoisotopic peak by Fourier transform ion cyclotron resonance mass spectrometry analysis. We then grew *Caenorhabditis elegans* cells in a medium containing living isotopically depleted *E. coli* cells, thereby producing the first isotopically depleted eukaryotic proteins. This is the first time isotopic depletion has been implemented for four isotopes ( $^1\text{H}$ ,  $^{12}\text{C}$ ,  $^{14}\text{N}$ , and  $^{16}\text{O}$ ), resulting in the highest degree of depletion ever used for protein analysis and further improving MS analysis.



## INTRODUCTION

The development of ultrahigh-resolution mass spectrometers has allowed for better characterization of large molecules, such as intact proteins, than was previously possible.<sup>1–3</sup> Accurate intact mass determination is one of the primary goals of mass spectrometry experiments; this is most easily achieved through the observation and accurate mass determination of the monoisotopic mass-to-charge ratio ( $m/z$ ) spectral peak.<sup>4</sup> Observation of the monoisotopic  $m/z$  peak is the simplest and most accurate way to determine monoisotopic mass, which is the mass of a molecule composed of the most abundant isotopes of each atom.

However, as molecular weight (MW) increases above approximately 1 kDa, other isotopic peaks become larger than the monoisotopic mass due to the contribution of the heavy isotopes of carbon, nitrogen, hydrogen, oxygen, and sulfur.<sup>5,6</sup> At masses greater than approximately 10 kDa, the monoisotopic  $m/z$  peak is of such low relative abundance that it is often undetectable or indistinguishable from noise.<sup>7</sup> Broadening of the isotopic distribution at increased molecular weight also results in decreased signal-to-noise ratio (SNR) because the detected signal is spread over a larger number of peaks, complicating

accurate intact mass determination and protein identification.<sup>8,9</sup>

The complexity of isotopic distributions and inability to accurately determine monoisotopic mass are considered a primary limitation in the identification and sequencing of intact proteins.<sup>10</sup> When there is no *a priori* knowledge of the protein(s) being analyzed, algorithms that model isotope distributions are employed to determine monoisotopic masses but often fail to accurately assign the monoisotopic  $m/z$  peak resulting in  $\pm 1$  or 2 Da mass errors.<sup>11–13</sup>

Almost three decades ago, a technique was developed to reduce the effects of isotopic distributions in mass spectrometry experiments. By culturing bacteria in isotopically depleted minimal media (i.e., inorganic salts, glucose depleted in  $^{13}\text{C}$ , and ammonium sulfate depleted in  $^{15}\text{N}$ ), bacterial cells were made to express FK506-binding protein that was depleted in heavy

**Received:** August 30, 2022  
**Revised:** December 21, 2022  
**Accepted:** December 22, 2022  
**Published:** January 19, 2023



isotopes of carbon and nitrogen.<sup>14</sup> The technique has been used to model the implications of increasing protein mass to observed SNR by Compton et al.<sup>4</sup> Gallagher et al. expanded on this idea by developing a method for the expression of isotopically depleted recombinant proteins depleted in <sup>13</sup>C and <sup>15</sup>N.<sup>15</sup> The benefit is that the relative abundance for the monoisotopic mass is increased so that it may be measured and observed directly in the mass spectrum. Additional depletion of oxygen and hydrogen should provide even greater enhancement.

In this study, we apply the technique of isotopic depletion to the analysis of proteins expressed by two organisms, *Escherichia coli* strain BL21 and *Caenorhabditis elegans*. *E. coli* BL21 cells were grown in M9 minimal medium formulated with both normal isotopic abundance and isotopically depleted nutrients. Unlike prior studies based solely on <sup>13</sup>C-depleted glucose and <sup>15</sup>N-depleted ammonium sulfate,<sup>15</sup> nutrients and water were additionally depleted in <sup>18</sup>O and <sup>2</sup>H, thereby achieving the highest-ever degree of isotopic depletion of proteins in a mass spectrometry study. Since heavy isotopes of C, H, O, and N were not fully depleted in the nutrients, the presence of some non-monoisotopic isotopologues is expected.

The media containing natural isotopic abundance and isotopically depleted *E. coli* were used to culture *C. elegans* N2 worms. The use of isotopically depleted media to culture *C. elegans* resulted in the production of isotopically depleted proteins, extending the isotopic depletion method to a eukaryotic organism for the first time. Both cell lysates were prefractionated by GELFrEE<sup>16</sup> prior to analysis by reversed-phase high-performance liquid chromatography (LC) tandem mass spectrometry (MS/MS) with collision-induced dissociation.<sup>16–18</sup> One of the main challenges is the lack of software for automated analysis of isotopically depleted proteins, but the manually validated examples discussed below serve to demonstrate the advantages of the technique.

## MATERIALS AND METHODS

**Media Preparation.** M9 minimum media were prepared as follows. Na<sub>2</sub>HPO<sub>4</sub>·2H<sub>2</sub>O, KH<sub>2</sub>PO<sub>4</sub>, NaCl, NH<sub>4</sub>Cl, MgSO<sub>4</sub>, CaCl<sub>2</sub> glucose, and <sup>15</sup>N-depleted (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> were obtained from Merck. <sup>13</sup>C-depleted glucose was purchased from Cambridge Isotope Laboratories. Deuterium-depleted water (DDW) containing ≤5 ppm D and 410 ppm <sup>18</sup>O was obtained from MTC Iceberg Ltd. (Moscow, Russia). M9 stock salt solutions (M9 5×SS) were prepared by dissolving 42.5 g of Na<sub>2</sub>HPO<sub>4</sub>·2H<sub>2</sub>O, 15.0 g of KH<sub>2</sub>PO<sub>4</sub>, and 2.5 g of NaCl in 1000 mL of either normal milli-Q water (for Normal M9 media) or DDW (for Depleted M9 media). Both solutions were autoclaved before proceeding to M9 media preparation. M9 media were prepared by mixing 800 mL of either normal milli-Q water (for Normal M9 media) or DDW (for Depleted M9 media), 200 mL of the corresponding M9 5×SS, 2.0 mL of 1 M MgSO<sub>4</sub>, 0.1 mL of 1 M CaCl<sub>2</sub>, 5 g of glucose, and 1 g of (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> (isotopically normal for Normal M9 media and <sup>13</sup>C-depleted glucose and <sup>15</sup>N-depleted (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> for Depleted M9 media). These media were filtered using 0.2 μm polyether sulfone (PES) filters (VWR) before use.

**Preparation of Isotopically Depleted *E. coli* Proteins.** *E. coli* BL21 cells were grown in 50 mL of M9 minimal media with natural-abundance and isotopically depleted nutrients at 37 °C using an incubation shaker (Ecotron, INFORS HT) at 200 rpm. After 36 h, *E. coli* cells were collected by centrifugation and resuspended in 5 mL of phosphate-buffered saline (PBS; Thermo Fisher Scientific) at pH 7.4, supplemented with

protease and phosphorylase inhibitor cocktail tablets (Roche). Resuspended cells were subjected to three freeze/thaw cycles and probe sonication; then, residual cell debris was removed by centrifugation for 15 min at 14,000 RCF. Protein concentrations in the resultant whole cell lysates were determined by Pierce microplate bicinchoninic acid assay (Thermo Fisher Scientific)<sup>19</sup> indicating protein concentrations of 537 μg/mL for natural abundance and 440 μg/mL for isotope depletion.

**Preparation of Isotopically Depleted *C. elegans* Proteins.** Natural-abundance and isotopically depleted *E. coli* BL21 proteins were cultured as described above. *C. elegans* N2 strains with mixed worm populations were grown on plates containing nematode growth medium (NGM) with *E. coli* OP50. The worms were collected and washed with M9 minimal medium and transferred into natural-abundance and isotopically depleted M9 minimal media. *E. coli* BL21 grown in natural M9 minimal media (as described above) were then used to inoculate normal and isotopically depleted M9 minimal media containing *C. elegans* at a 1:50 dilution. *C. elegans* and the bacteria were cocultured in 50 mL of media at 25 °C. The coculture was performed in an incubation shaker at 100 rpm. After 72 h, the *C. elegans* were separated from bacteria by keeping the mixture still for 20 min at room temperature, removing bacteria in the upper layer, and then the worm pellets were washed with M9 minimal medium. The keeping still/washing cycles were repeated three times before the worms were resuspended in 1.5 mL of PBS supplemented with protease and phosphatase inhibitors. After sonication and centrifugation, protein concentrations were measured by Pierce microplate bicinchoninic acid assay. The protein concentrations of the whole-worm extracts were 1.4 mg/mL for the natural-abundance extract and 3.3 mg/mL for the isotopically depleted extract.

**Prefractionation.** The normal and isotopically depleted samples were desalted and exchanged into 200 μL of 1% w/v sodium dodecyl sulfate solution by use of Amicon Ultra 3K molecular weight cutoff centrifugal filters. Initial salt concentration (150 mM) was decreased by a factor of ~10. Each sample was split into two 100 μL aliquots to which 20 μL of water and 30 μL of 5× Tris-acetate GELFrEE sample buffer were added, followed by gentle vortexing and centrifugation at 16,000 RCF for 10 min to pellet any remaining debris. A GELFrEE 8100 fractionation system (Expedeon) employing a 10% Tris-acetate cartridge was used for MW-based separation according to manufacturer protocols.<sup>20</sup> Natural-abundance and isotopically depleted samples were loaded in technical duplicates (i.e., two lanes per sample) with 225 μg of *E. coli* or *C. elegans* protein in each lane. In total, 12 fractions were collected from each lane.

Protein from each fraction was separated and visualized by SDS-PAGE using Mini-PROTEAN TGX precast gels (Bio-Rad) according to the manufacturer's recommended protocol and visualized by silver staining with a Pierce silver stain kit (Thermo Fisher Scientific) to evaluate relative protein content and the MW range of proteins in each fraction.

After GELFrEE separation and prior to LC-MS/MS analysis, each fraction underwent methanol–chloroform–water precipitation to remove detergents and contaminants.<sup>21</sup> Precipitated protein was reconstituted in 35 μL of solvent A (94.7% water, 5% acetonitrile, and 0.3% formic acid) prior to LC separation.

**Liquid Chromatography Tandem Mass Spectrometry.** Two sets of reversed-phase chromatography trap and analytical columns were packed in fused silica microcapillaries (Trap, 360 μm o.d. × 150 μm i.d.; Analytical, 360 μm o.d. × 75 μm i.d.), for natural-abundance and isotopically depleted experiments. Trap

columns were made with the packed bed 2–3 cm in length and analytical columns packed to 15–17 cm length. Columns were packed (length given) with PLRP-S packing material [macro-porous poly(styrene-divinylbenzene); 5  $\mu\text{m}$  particles, 1000  $\text{\AA}$  pore, Agilent Technologies, Palo Alto, CA]. The fabricated columns were connected directly to the nano-HPLC system (ACQUITY M-class, Waters, Milford, MA). Acetic acid ( $\geq 99.99\%$ ) and LC-MS grade isopropyl alcohol were purchased from Sigma-Aldrich (St. Louis, MO). LC-MS grade water, acetonitrile, methanol, and chloroform were purchased from Honeywell Burdick and Jackson (Muskegon, MI). Samples were analyzed in duplicate with 4  $\mu\text{L}$  injections at a 2.5  $\mu\text{L}/\text{min}$  flow rate for loading/trapping and 0.3  $\mu\text{L}/\text{min}$  for the gradient separation. Solvent A (0.3% formic acid in 5% acetonitrile and 94.7% water) and Solvent B (0.3% formic acid in 4.7% water, 47.5% acetonitrile, and 47.5% isopropyl alcohol) were used as the mobile phase with the %B increasing over the gradient time of 60 min from 5% to 80%. Analytical gradients were adjusted to fractions based on the expected MW range, increasing in maximum %B for higher MW fractions due to the trend of increasing hydrophobicity with increasing protein MW.

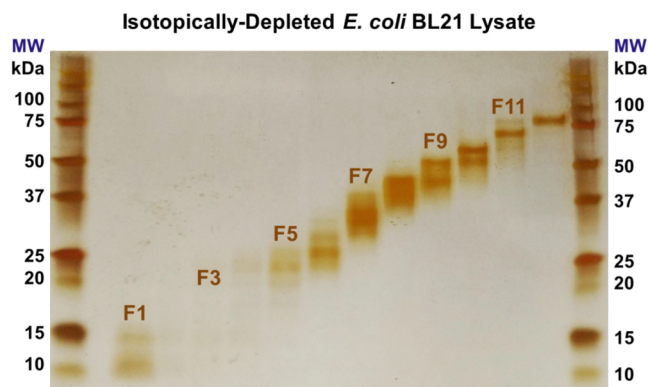
Electrospray ionization was performed with a source voltage of +3.5 kV applied directly to a 15  $\mu\text{m}$  fused-silica PicoTip emitter (New Objective, Woburn, MA) packed with 3 mm PLRP-S resin. The heated metal capillary temperature was 325  $^{\circ}\text{C}$ . All mass spectrometry experiments were performed at the National High Magnetic Field Laboratory (Tallahassee, FL) with a hybrid 21 T FT-ICR mass spectrometer, which includes a front-end Velos Pro linear RF ion trap (Thermo Fisher Scientific, San Jose, CA).<sup>22</sup> All intact protein mass spectra ranged from  $m/z$  300 to 2000 with a 1E6 automatic gain control (AGC) target and were recorded as 4 summed transient acquisitions. Data-dependent acquisition (DDA) was performed using Xcalibur (Thermo Fisher Scientific) software and employed CID fragmentation (10 ms; normalized collision energy 35%) with a 2E5 AGC target and single transient acquisition. DDA MS/MS was performed for the top 2 highest magnitude peaks in the intact mass scan in the  $m/z$  300–2000 range. Prior to FT-ICR MS analysis, multiple accumulations of fragment ions were stored in an external multipole storage device to build larger ion populations and improve spectral SNR.<sup>23</sup> All mass spectra were collected with a 0.76 s transient (resolving power 103 K at  $m/z$  1211).

**Data Analysis.** Raw mass spectral data from natural-isotope-abundance *E. coli* and *C. elegans* proteins were submitted to the National Resource for Translational and Developmental Proteomics (Northwestern University, Evanston, IL) for processing with the TDPortal high-performance computing environment at Northwestern University (available for academic collaborators here: <http://nrtdp.northwestern.edu/tdportal-request/>). TDPortal uses a database of known proteoform sequences to identify them in normal isotopic abundance following  $m/z$ -to- $m$  deconvolution of raw data with the Xtract algorithm (Thermo Fisher Scientific). Then the same protein was manually identified in the corresponding isotopically depleted raw data with consideration given to retention time, fraction number, fragment ion assignments, and calculated intact monoisotopic mass. This same method was implemented for the *C. elegans* raw data. All data were manually interpreted, and an SNR threshold of 5 was used for peak assignment at the MS1 and MS2 levels. For MS2 data analysis, a 10 ppm difference between measured mass and mass computed from the assigned elemental composition was accepted if the fragment isotopic distribution

was observed or one of 5 ppm if the distribution was not observed (monoisotopic  $m/z$  peak only). Custom software was written to model the isotopic distributions of expected intact protein and fragment ions and to calculate isotopologue masses based on user-defined amino acid sequence composition, accurate mass, and isotope abundances to enable confident assignment of isotopically depleted ions. ProSight Lite was used to generate sequence coverage maps that show backbone amide bond cleavages have been identified.<sup>24</sup> IsoPro 3.1 (available at <http://sites.google.com/site/isoproms/>) was used for isotope abundance and degree of depletion calculations and SciDAVIS for data visualization.<sup>25</sup>

## RESULTS AND DISCUSSION

***E. coli* Protein Analysis.** Due to the compositional complexity of proteins expressed by cells, a prefractionation step is recommended. GELFrEE MW separation followed by SDS-PAGE with silver staining provided visualization of the MW ranges for each fraction. We observed lower protein concentrations in fractions 1–4 for both natural-isotope-abundance and isotopically depleted samples, as seen in Figure 1 for *E. coli*. Prior to LC-MS analysis, fractions 1 and 2 were



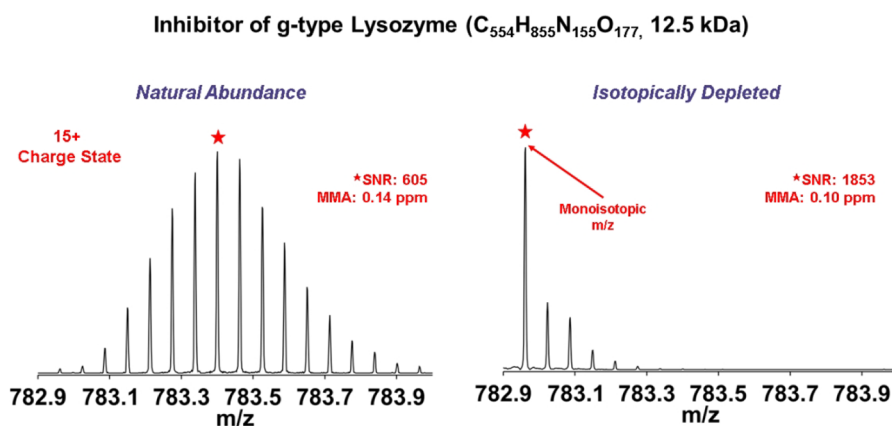
**Figure 1.** A picture of SDS-PAGE molecular weight separation gel stained with silver nitrate shows each fraction collected from *E. coli* GELFrEE fractionation (F#) and their corresponding molecular weight ranges (MW scale on either side for control). The contrast of each stained band gives a sense of relative concentration in each fraction. This example is one of four GELFrEE fractionations performed.

combined as well as fractions 3 and 4. The rest of the fractions were analyzed individually. No GELFrEE fractions were combined for the *C. elegans* samples. Further, the insight into the MW ranges for each fraction allowed for experimental parameters to be tailored to account for a large dynamic MW range of proteins.

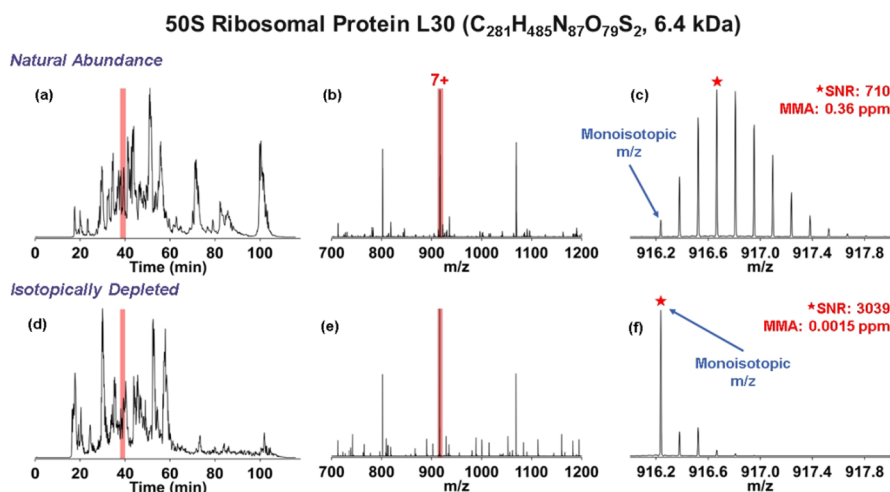
SNR and mass measurement accuracy (MMA) are the two figures of merit that improve by use of isotopic depletion. In Figure 2, 15+ ions of inhibitor of g-type lysozyme show the contrast between the two isotopic distributions. The abundant monoisotopic peak in the isotopically depleted samples allows for easy calculation of the monoisotopic mass, which is the most accurate mass characteristic. For SNR and MMA calculations, the highest-abundance charge state ion is chosen due to potential spectral overlaps, and low SNR of the monoisotopic peak limited the accuracy and reliability of *E. coli* protein mass determination throughout the  $m/z$  range.<sup>7</sup>

As an example of various manually analyzed *E. coli* proteins, intact protein mass spectra of 50S ribosomal protein L30 are

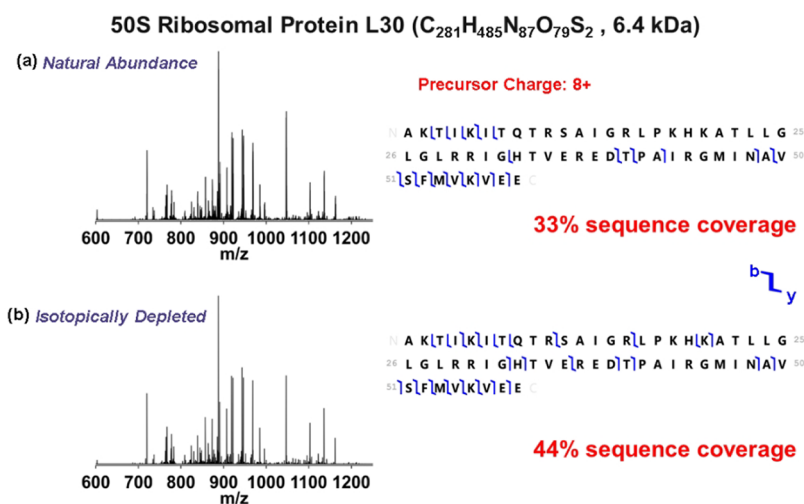




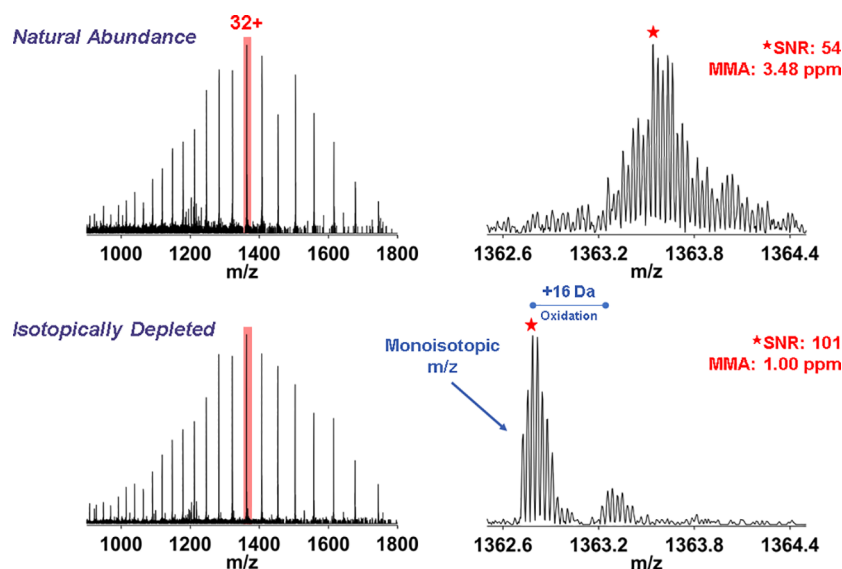
**Figure 2.** (+) ESI 21 T FT-ICR mass spectra for natural isotope abundance (left) along with the new isotopically depleted data (right) for the inhibitor of g-type lysozyme protein. These two spectra are zoom-insets of the 15+ charge state; the abundant monoisotopic  $m/z$  in the isotopically depleted samples allows for easy calculation of monoisotopic mass. Low SNR and spectral overlap hinder the identification of the monoisotopic peak in the natural-abundance data.



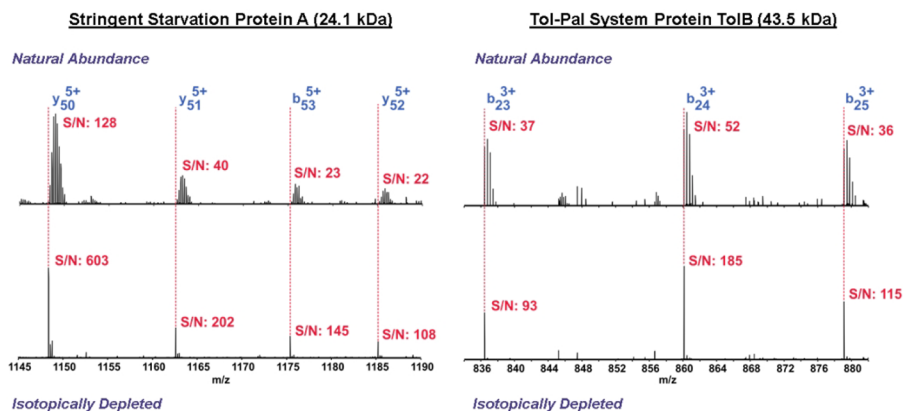
**Figure 3.** An example of *E. coli* protein, 50S ribosomal protein L30, is 6.4 kDa in size and contains 281 carbons. (a) Natural-abundance fraction chromatogram with the retention time highlighted for the protein. (b) Charge state distribution of the protein with the 7+ charge state highlighted. (c) Zoom-in of the 7+ charge state. (d–f) The bottom panels share the same information for the isotopically depleted sample. The monoisotopic peaks are labeled with the S/N and MMA for the most abundant peak included.



**Figure 4.** The two spectra on the left are for the CID fragmentation of the 8+ precursor for (a) natural abundance and (b) isotopically depleted. Blue vertical lines indicate cleavages with the flags on the top (b ion) and bottom (y ion) indicating fragment ion type.



**Figure 5.** This example is of a protein that is 43.5 kDa in size. The intact MS (left) is followed by a zoom-in for the 32+ charge state (right). S/N and MMA for the highest magnitude peak in each distribution. In the zoom-in of the isotopically depleted data, the monoisotopic peak is clearly detected. In addition, we see the addition of 16 Da which indicates oxidation.



**Figure 6.** Two *E. coli* proteins are used to compare S/N improvement seen at the fragment ion level. Ion type, charge, and position on the backbone are labeled at the highest magnitude natural-abundance monoisotopic peak. The bottom spectra are for the isotopically depleted data with S/N of fragment monoisotopic peaks labeled.

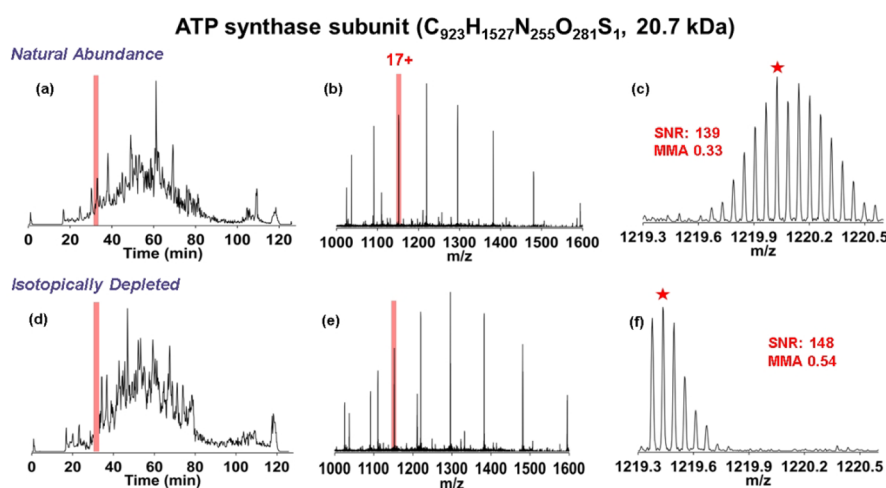
shown in Figure 3 along with a direct comparison of total ion chromatograms. A 4.2-fold improvement in SNR and a 240-fold improvement in MMA are observed for the 7+ molecular ions of the isotopically depleted protein relative to the natural-abundance protein. The elution times for the natural isotopic abundance and isotopically depleted proteins differ slightly, most likely due to the use of two separate sets of columns and differences in the concentration of the two protein samples. In addition, the reduction in number of isotopologues and confident assignment of monoisotopic peaks resulted in an increase in sequence coverage by 11%, as seen in Figure 4. Various proteins were identified in both sample types for comparison of chromatographic retention times, intact protein mass spectra, CID fragmentation spectra, SNR, and MMA.

As noted above, identification of proteins becomes more difficult with increasing MW. For example, for Tol-pal system protein (TolB), which is 7 times more massive than the 50S ribosomal protein L30, we start to see the role of isotopologues in the determination of the monoisotopic peak (Figure 5). In the natural isotopic abundance data, no monoisotopic mass could be assigned directly due to low SNR and spectral overlap of a

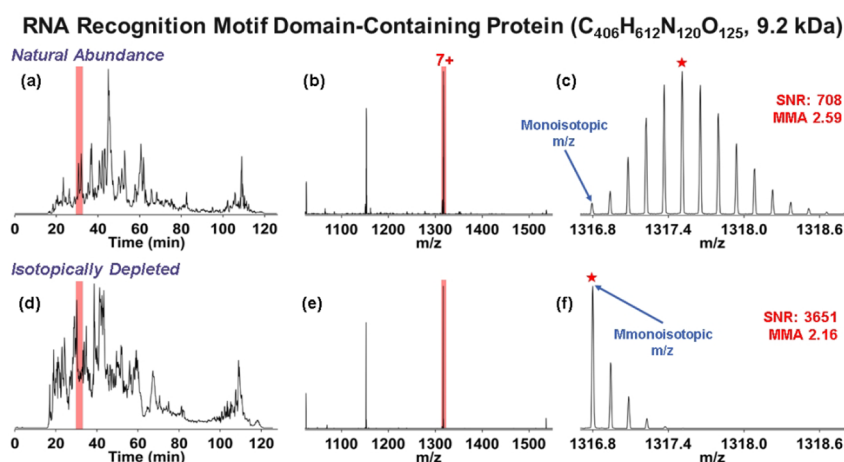
neighboring distribution. In contrast, for the isotopically depleted data we could not only clearly identify the monoisotopic  $m/z$  but also resolve the oxidized satellite. In the isotopically depleted data, we observe the SNR doubled and the MMA improved by 3-fold. As previously mentioned, one of the hurdles of intact protein analysis is widening isotopic distributions which hinder our ability to obtain accurate intact mass. In Figure 5, we can clearly see the improvement gained by isotopic depletion.

We observed SNR improvement not only at the intact MS1 level but also in the MS2 fragment mass spectra, as seen in Figure 6. Fragmentation spectra suffer in general from a significant decrease in SNR, which limits the sequencing information. For stringent starvation protein A, an  $m/z$ -scale expansion for four manually identified  $y_{50-53}$  5+ ions and the precursor spectra showed SNR improvement up to 6.3-fold.

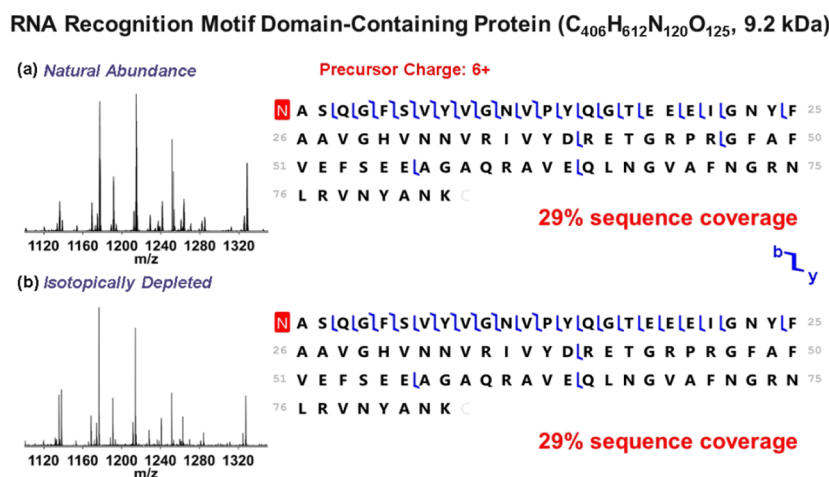
**C. elegans Protein Analysis.** The same analysis parameters were used to analyze a handful of protein examples in the *C. elegans* data. One such example, ATP synthase subunit, can be seen in Figure 7. A slight improvement in both the SNR and MMA is observed in the zoom-in of the 17+ charge state. The



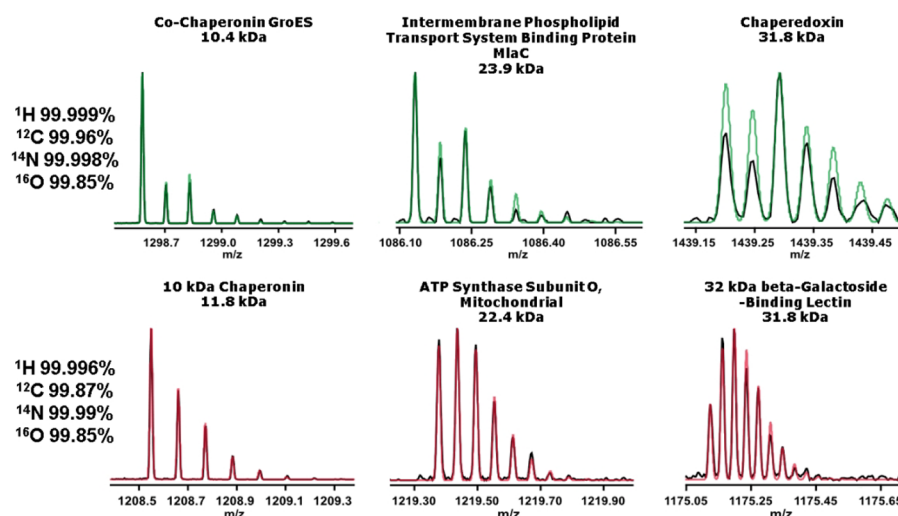
**Figure 7.** LC and MS1 comparisons of a *C. elegans* protein, ATP synthase subunit, a 20.7 kDa protein. (a) Fraction chromatogram with the retention time highlighted for the protein in the natural-abundance sample. (b) Charge state distribution of the protein with the 17+ charge state highlighted. (c) Zoom-in of the 17+ charge state. (d–f) The bottom panels share the same information for the isotopically depleted sample. The monoisotopic peaks are labeled with the SNR and MMA for the highest magnitude peak included.



**Figure 8.** An example of one *C. elegans* protein identification, RNA recognition motif domain-containing protein, is a 9.2 kDa protein. (a) Fraction chromatogram with the retention time highlighted for the protein in the natural-abundance sample. (b) Charge state distribution of the protein with the 7+ charge state highlighted. (c) Zoom-in of the 7+ charge state. (d–f) The bottom panels share the same information for the isotopically depleted sample. The monoisotopic peaks are labeled with the S/N and MMA for the highest magnitude peak included.



**Figure 9.** The two spectra on the left are of the CID fragmentation of the 6+ precursor for (a) natural abundance and (b) isotopically depleted. Blue vertical lines indicate cleavages with the flags on the top (b ion) and bottom (y ion) indicating fragment ion type. No sequence coverage improvement is present here, much like many proteins manually identified in the *C. elegans* samples.



**Figure 10.** On top are three *E. coli* proteins that were identified with varying masses. The same was done for three *C. elegans* proteins (bottom), with the raw data (black) overlapped by the estimated isotope abundances overlapping (green and red).

monoisotopic peak cannot be clearly distinguished from the noise in this example. RNA recognition motif domain-containing protein is shown in Figure 8. As for the *E. coli* data, the chromatographic traces show high similarity but are not identical. The MS1 spectra contain a charge state distribution with the most abundant 7+ charge state highlighted; it was used for the  $m/z$  scale expansion comparison. For the isotopically depleted protein, a 5.2-fold increase in SNR was observed, but the MMA was not improved. The MS2 data for this protein can be seen in Figure 9. There was no improvement in the sequence coverage due to isotopic depletion, but 29% sequence coverage is still a step forward from most published experiments for proteins of this size from single scan data analysis.

**Depletion Estimates and Modeling.** Data analysis for proteins showed a lower degree of isotopic depletion for *C. elegans* than for *E. coli*. With the knowledge of nutrient depletion, estimates of the degree of depletion in proteins were modeled for best fit. The abundances of  $^1\text{H}$ ,  $^{12}\text{C}$ ,  $^{14}\text{N}$ , and  $^{16}\text{O}$  for each case are shown along with the raw data (black) overlapped with the modeled data (green and red). The depletion modeling of *E. coli* protein data showed that all isotopes were depleted to various degrees:  $^1\text{H}$  99.999%,  $^{12}\text{C}$  99.96%,  $^{14}\text{N}$  99.998%, and  $^{16}\text{O}$  99.85%. For the *C. elegans* data we observed a slight decrease in depletion of each isotope:  $^1\text{H}$  99.996%,  $^{12}\text{C}$  99.87%,  $^{14}\text{N}$  99.99%, and  $^{16}\text{O}$  99.85%. To test the estimated isotopic abundances, multiple proteins in the MW range of 10–31.8 kDa were modeled. The estimated abundances from modeling (red and green) are overlaid on the raw data (black) in Figure 10. There is a limitation to the accuracy of the isotopic abundances that we can determine using the above approach. As proteins increase in MW, the inaccuracy of the modeled isotopic abundances results in larger deviations between the observed and theoretical spectra. Moreover, we can see the impact the small decrease in depletion has on the isotopic distribution when we compare proteins of relatively the same MW. This decrease is potentially why the SNR and MMA improvement is greater for the *E. coli* data than for the *C. elegans* data.

## CONCLUSIONS

The above results provide potential solutions to two of the main challenges for identification of intact proteins: identification of the monoisotopic peak and adequate SNR. Although the

improvements in S/N and MMA for the isotopically depleted *C. elegans* data were not as dramatic as for *E. coli*, they demonstrated successful propagation of isotopic depletion along the food chain.

Processing of the raw data from the natural-isotope-abundance samples yielded identification of many-fold more proteins than the isotopically depleted raw data, in spite of higher MS1 and MS2 spectra quality for the latter. Upon further inspection, the data analysis software was determined to be the problem. The currently available software uses an isotopic distribution modeling algorithm based on natural isotopic abundances, averagine, and thus cannot accurately detect and assign isotopically depleted distributions.<sup>11</sup> As of today, there is no published software that can analyze isotopically depleted proteins.

Successful isotopic depletion of *E. coli* and *C. elegans* proteins gives hope for extending the technique to proteins derived from mammalian cells grown in isotopically depleted media based, e.g., on hydrolysis products of bacterial or *C. elegans* proteins.

## AUTHOR INFORMATION

### Corresponding Author

Alan G. Marshall – National High Magnetic Field Laboratory, Tallahassee, Florida 32310, United States; Department of Chemistry and Biochemistry, Florida State University, Tallahassee, Florida 32306, United States; [orcid.org/0000-0001-9375-2532](https://orcid.org/0000-0001-9375-2532); Phone: +1 850-644-0529; Email: [marshall@magnet.fsu.edu](mailto:marshall@magnet.fsu.edu); Fax: +1 850-644-1366

### Authors

Zeljka Popovic – National High Magnetic Field Laboratory, Tallahassee, Florida 32310, United States; Department of Chemistry and Biochemistry, Florida State University, Tallahassee, Florida 32306, United States; [orcid.org/0000-0003-2411-6238](https://orcid.org/0000-0003-2411-6238)

Lissa C. Anderson – National High Magnetic Field Laboratory, Tallahassee, Florida 32310, United States; [orcid.org/0000-0001-8633-0251](https://orcid.org/0000-0001-8633-0251)

Xuepei Zhang – Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Solna 171 77 Stockholm, Sweden



David S. Butcher – National High Magnetic Field Laboratory, Tallahassee, Florida 32310, United States

Greg T. Blakney – National High Magnetic Field Laboratory, Tallahassee, Florida 32310, United States; [orcid.org/0000-0002-4205-9866](https://orcid.org/0000-0002-4205-9866)

Roman A. Zubarev – Department of Medical Biochemistry and Biophysics, Karolinska Institutet, Solna 171 77 Stockholm, Sweden; [orcid.org/0000-0001-9839-2089](https://orcid.org/0000-0001-9839-2089)

Complete contact information is available at:  
<https://pubs.acs.org/10.1021/jasms.2c00242>

## Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

A portion of this work was performed at the National High Magnetic Field Laboratory ICR User Facility, which is supported by the National Science Foundation Division of Chemistry through Cooperative Agreement No. DMR-1644779 and the State of Florida. Another portion was supported by the EU FET OPEN project TopSpec and the Swedish Research Council project 2017-04303.

## REFERENCES

- (1) Yergey, J.; Heller, D.; Hansen, G.; Cotter, R. J.; Fenselau, C. Isotopic Distributions in Mass Spectra of Large Molecules. *Anal. Chem.* **1983**, *55* (2), 353–356.
- (2) Domon, B.; Aebersold, R. Mass Spectrometry and Protein Analysis. *Science (1979)* **2006**, *312* (5771), 212–217.
- (3) Mann, M.; Hendrickson, R. C.; Pandey, A. Analysis of Proteins and Proteomes by Mass Spectrometry. *Annu. Rev. Biochem.* **2001**, *70* (1), 437–473.
- (4) Compton, P. D.; Zamdborg, L.; Thomas, P. M.; Kelleher, N. L. On the Scalability and Requirements of Whole Protein Mass Spectrometry. *Anal. Chem.* **2011**, *83* (17), 6868–6874.
- (5) Yergey, J.; Heller, D.; Hansen, G.; Cotter, R. J.; Fenselau, C. Isotopic Distributions in Mass Spectra of Large Molecules. *Anal. Chem.* **1983**, *55* (2), 353–356.
- (6) Zubarev, R. A.; Bonddarenko, P. v. An A Priori Relationship between the Average and Monoisotopic Masses of Peptides and Oligonucleotides. *Rapid Commun. Mass Spectrom.* **1991**, *5* (6), 276–277.
- (7) Zubarev, R. A.; Demirev, P. A.; Haakansson, Per.; Sundqvist, B. U. R. Approaches and Limits for Accurate Mass Characterization of Large Biomolecules. *Anal. Chem.* **1995**, *67* (20), 3793–3798.
- (8) Valkenborg, D.; Mertens, I.; Lemiere, F.; Witters, E.; Burzykowski, T. The Isotopic Distribution Conundrum. *Mass Spectrom Rev.* **2012**, *31*, 96–109.
- (9) Lubec, G.; Afjehi-Sadat, L. Limitations and Pitfalls in Protein Identification by Mass Spectrometry. *Chem. Rev.* **2007**, *107* (8), 3568–3584.
- (10) Ahlf, D. R.; Thomas, P. M.; Kelleher, N. L. Developing Top down Proteomics to Maximize Proteome and Sequence Coverage from Cells and Tissues. *Curr. Opin. Chem. Biol.* **2013**, *17* (5), 787–794.
- (11) Senko, M. W.; Beu, S. C.; McLafferty, F. W. Determination of Monoisotopic Masses and Ion Populations for Large Biomolecules from Resolved Isotopic Distributions. *J. Am. Soc. Mass Spectrom.* **1995**, *6* (4), 229–233.
- (12) Schaffer, L. v.; Millikin, R. J.; Miller, R. M.; Anderson, L. C.; Fellers, R. T.; Ge, Y.; Kelleher, N. L.; LeDuc, R. D.; Liu, X.; Payne, S. H.; Sun, L.; Thomas, P. M.; Tucholski, T.; Wang, Z.; Wu, S.; Wu, Z.; Yu, D.; Shortreed, M. R.; Smith, L. M. Identification and Quantification of Proteoforms by Mass Spectrometry. *Proteomics* **2019**, *19* (10), 1800361.
- (13) Zubarev, R. A.; Demirev, P. A. Isotope Depletion of Large Biomolecules: Implications for Molecular Mass Measurements. *J. Am. Soc. Mass Spectrom.* **1998**, *9* (2), 149–156.
- (14) Marshall, A. G.; Senko, M. W.; Li, W.; Li, M.; Dillon, S.; Guan, S.; Logan, T. M. Protein Molecular Mass to 1 Da by  $^{13}\text{C}$ ,  $^{15}\text{N}$  Double-Depletion and FT-ICR Mass Spectrometry. *J. Am. Chem. Soc.* **1997**, *119* (2), 433–434.
- (15) Gallagher, K. J.; Palasser, M.; Hughes, S.; Mackay, C. L.; Kilgour, D. P. A.; Clarke, D. J. Isotope Depletion Mass Spectrometry (ID-MS) for Accurate Mass Determination and Improved Top-Down Sequence Coverage of Intact Proteins. *J. Am. Soc. Mass Spectrom.* **2020**, *31* (3), 700–710.
- (16) Tran, J. C.; Doucette, A. A. Gel-Eluted Liquid Fraction Entrapment Electrophoresis: An Electrophoretic Method for Broad Molecular Weight Range Proteome Separation. *Anal. Chem.* **2008**, *80* (5), 1568–1573.
- (17) Shapiro, A. L.; Vinuela, E.; Maizel, J. V. Molecular Weight Estimation of Polypeptide Chains by Electrophoresis in SDS-Polyacrylamide Gels. *Biochem. Biophys. Res. Commun.* **1967**, *28* (5), 815–820.
- (18) Banerjee, R.; Maheswarappa, N. B.; Dasoju, S.; Ande, S. S. OFFGEL and GELFrEE Fractionation: Novel Liquid-Phase Protein Recovery Strategies in Proteomics Studies. *TrAC Trends in Analytical Chemistry* **2021**, *140*, 116282.
- (19) Walker, J. M. *Bicinchoninic Acid (BCA) Assay for Protein Quantitation* **2009**, 11–15.
- (20) Witkowski, C.; Harkins, J. Using the GELFREE 8100 Fractionation System for Molecular Weight-Based Fractionation with Liquid Phase Recovery. *J. Visualized Exp.* **2009**, *34*, 1–2.
- (21) Wessel, D.; Flügg, U. I. A Method for the Quantitative Recovery of Protein in Dilute Solution in the Presence of Detergents and Lipids. *Anal. Biochem.* **1984**, *138* (1), 141–143.
- (22) Hendrickson, C. L.; Quinn, J. P.; Kaiser, N. K.; Smith, D. F.; Blakney, G. T.; Chen, T.; Marshall, A. G.; Weisbrod, C. R.; Beu, S. C. 21 T Fourier Transform Ion Cyclotron Resonance Mass Spectrometer: A National Resource for Ultrahigh Resolution Mass Analysis. *J. Am. Soc. Mass Spectrom.* **2015**, *26* (9), 1626–1632.
- (23) Schaub, T. M.; Hendrickson, C. L.; Horning, S.; Quinn, J. P.; Senko, M. W.; Marshall, A. G. High-Performance Mass Spectrometry: Fourier Transform Ion Cyclotron Resonance at 14.5 T. *Anal. Chem.* **2008**, *80* (11), 3985–3990.
- (24) Fellers, R. T.; Greer, J. B.; Early, B. P.; Yu, X.; LeDuc, R. D.; Kelleher, N. L.; Thomas, P. M. ProSight Lite: Graphical Software to Analyze Top-down Mass Spectrometry Data. *Proteomics* **2015**, *15* (7), 1235–1238.
- (25) Benkert, T.; Franke, K.; Garriga, M.; Narayanankutty, A.; Standish, R. *SciDAVis*, 2007. <https://github.com/SciDAVis>.